A Novel Approach to EEG Neurofeedback via Reinforcement Learning

Aman Bhargava, Kyle O'Shaughnessy, Steve Mann
MannLab Canada, 330 Dundas Street West, Toronto, Ontario, M5T 1G5

***Abstract*—Since the invention of EEG brain scanning technology, cognitive response modeling and brain state optimization has been a topic of great interest and value. In particular, applications for improving musical therapy via neurofeedback have shown promise for brain state optimization.**

Here, we propose a novel Humanistically Intelligent (HI) system where brain waves are interpreted by a real-time deep reinforcement learning agent that controls an audio modulation system in order for the user to achieve a target brain state. The modulated audio is then visualized using a simulated Sequential Wave Imprinting Machine (SWIM). In our tests comparing the proposed system to a conventional neurofeedback system, we found that the proposed system consistently led to a more meditative brain state with p = 0.06.

We show that the proposed system is promising for brain state optimization tasks, advancing the intelligent utilization of EEG brain scan data in Humanistically Intelligent feedback loops.

***Index Terms*—brain-computer interface (BCI), machine learning, reinforcement learning, humanistic intelligence (HI), wearable technology, cognitive science, music therapy.**

I. Introduction

Brain-computer interface technology has long held the promise of enhancing our understanding of the mind's response to stimulus [1]. Such an understanding would enable the creation of feedback systems that can assist humans in achieving arbitrary brain states, including increased focus, mindfulness, and attention. Currently, such feedback systems far from perfect — much of the information potentially contained in the data is uninterpretable or underutilized in current methods [2].

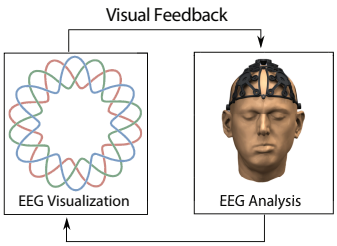

Fig. 1: A conventional neurofeedback loop, prototyped in *Electrical Engineering Design with the Subconscious Mind* [3]. EEG data directly controls a stimulus that is directly fed back to the user.

One paradigm that has shown much promise is neurofeedback (Fig. 1). Brain scan information is conveyed directly back to the user through audio/visual channels [4]. Studies have confirmed that neurofeedback positively affects mental state [5], but the system fails to address how different individuals may benefit from unique audio/visual feedback experiences.

The question remains: how should one optimally interpret and utilize brain waves in order to improve user's brain state? In the proposed system, EEG data is passed first to a deep reinforcement learning agent that processes the information with a user-specified reward function (e.g. correlates of meditation in EEG signal). The deep reinforcement learning agent controls an audio modulation system with the goal of maximize the reward function, thus optimizing brain state.

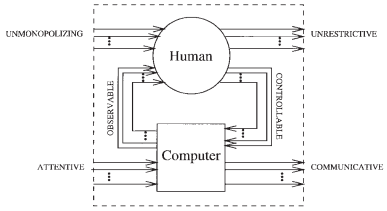

Fig. 2: Signal flow paths for a Humanistically Intelligent control system. For the purposes of the proposed system, the control is exerted via EEG signals while the computer's behavior is observed via auditory and visual feedback. Though the current implementation is for meditation enhancement, there is no fundamental monopolizing or restriction on the user's actions or behavior.

This system is an advanced form of a Humanistically Intelligent (HI) feedback loop [6], [7]. Rather than creating a non-personalized, non-adaptive experience for the user, we utilize the human mind and a deep reinforcement learning agent as a parts of the feedback loop to improve its effectiveness.

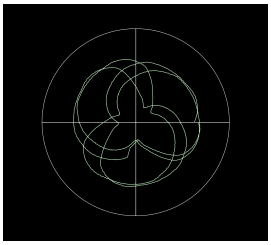

Fig. 3: Rotary SWIM Simulation in action, visualizing the waveform of a Tibetan singing bowl.

Simulated Sequential Wave Imprinting (SWIM) technology is used to provide a visual component to the feedback loop. SWIM utilizes multi-mediated reality to make invisible physical phenomenology visible [8]. It can thus be used to visualize

the sound wave phenomena produced by the proposed system. Analogue SWIM is composed of an array of light sources that receives input from a computing device that receives and processes wave information in real time. When the SWIM is moved through space, the phenomenology is made visible [9]. Past works have indicated the strong potential for SWIM as a meditative tool [3].

For the purposes of this experiment, the reward function of the reinforcement learning system was set to promote a more meditate brain state. The null and alternative hypothesis were as follows:

H_0: There is no difference between brain state improvement for individuals using the proposed system compared to those using a conventional neurofeedback system.

H_1: The proposed system causes a greater improvement in brain state than the conventional neurofeedback system.

II. Materials and Methodology

A. The Proposed System

For those unfamiliar with reinforcement learning, a brief overview of the underlying concepts for the reinforcement learning techniques used can be found in Section II B.

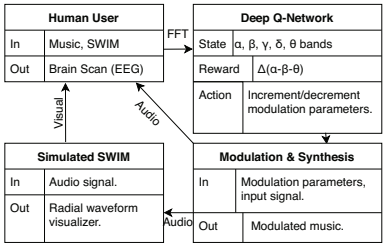

Fig. 4: Block diagram of the proposed system. The user's EEG is processed into frequency bands and streamed through the Deep Q-Network where an action is processed each second. This action changes the audio modulation parameters, and the processed signal is observed by the user via auditory and visual input using a Simulated SWIM.

For the proposed system, brain scan data was collected via the Muse 2 EEG headset. Data was streamed to a custom Bluetooth Low-Energy server at 256 samples per second where a windowed Fast Fourier Transform (FFT) algorithm was applied to the last 512 samples each second. The power spectral distribution was then calculated and averaged into alpha (8-15 Hz), beta (16-31 Hz), theta (4-7 Hz), and delta (¡4 Hz) frequency bands [1] [2].

The reward for the Deep Q-Network was calculated as follows in order to reflect the user's meditative state [10] based on their AF7 electrode potential.

$$R = \alpha - \beta - \theta$$

Meanwhile, the state space for the reinforcement learning agent was composed of the frequency bands defined above. The action space consisted of either incrementing or decrementing the singular audio modulation parameter. Memory replay [11] and exploration decay [12] were employed in order to increase the training speed and efficacy of the Deep Q-Network.

The audio modulation system simply applied a sinusoidal low-frequency oscillator (LFO) to the gain of the audio where the frequency was supplied by the deep reinforcement learning system. Only one parameter was chosen due to the constraint on training data. Since training restarted for each new test subject and only lasted 600 seconds (600 data points total), it was most feasible for the algorithm to learn the user's response to changes in only 1 parameter based on their brain state.

For the purposes of these experiments, sustained synthesized chords were used as input to the modulation system. The audio was then visualized by the simulated SWIM.

Fig. 5: The proposed system in action. EEG data is streamed from the Muse headset to be processed on the BLE server. Processed audio is visualized through simulated SWIM as the user attempts to meditate.

B. Q-Learning and Deep Q-Network Background

The Q-learning algorithm attempts to predict the expected value of any action a given a state s assuming that the action with the maximum expected value from the Q-learning algorithm is taken at every subsequent values [13]. These expected values are known as Q values. For conventional Q-Learning, a function is learned such that the following Bellman Optimality condition is met [14]:

$$Q(s,a) = R_{t+1} + \gamma \max_{a'} Q(s',a')$$

Where $Q(s,a)$ is the Q function that returns the maximum expected value of taking action a given state s and R_{t+1} is the reward reaped at the very next time step due to action a. s' and a' represent the subsequent state and action options. γ is the discount factor for far future rewards.

When this condition is met, following the action that maximizes Q at every step would lead to the maximum net reward possible. Since there are no ground truth values to train the Q-function on, values are updated as follows based on reward information:

$$Q(s_t,a_t) \leftarrow Q(s_t,a_t) + \alpha[R_{t+1} + \gamma \max_{a'}(Q(s',a') - Q(s_t,a_t))]$$

Where α is the learning rate.

For Deep Q-Networks (DQN), the Q-function is approximated by a neural network. The network takes in a state vector s and outputs the Q-values for each potential action as follows [15]:

$$DQN(s) = \{Q(s, a_1), Q(s, a_2), ..., Q(s, a_n)\}$$

The network is trained using the right-hand side of the above update equation to approximate ground truths to back-propagate based on reward/action/state tuples.

C. Neurofeedback System

The conventional neurofeedback (Fig. 1) system consists of much the same components as outlined above for the proposed system (Fig. 4) and has shown promise for brain state optimization in therapy and meditation [4], [5]. The primary difference between the neurofeedback system and the proposed system was that the Reinforcement Learning system was bypassed and the meditation score was used to directly modulate the low frequency gain oscillator in the neurofeedback system.

III. Results and Discussion

A. Results

For each of the participants in the trials, they were asked to meditate by focusing on their breath while wearing the EEG apparatus connected to the proposed system and the conventional neurofeedback systems for 10 minutes each (order was randomized for each participant). As they observed the modulated audio and the SWIM visualization, their brain activity was recorded along with the modulation parameters sent to the audio system (Fig. 6).

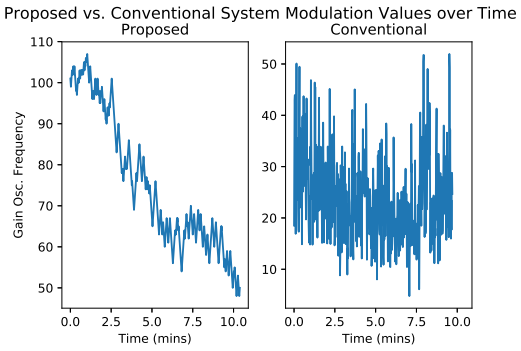

Fig. 6: Modulation value over time from the proposed system and the conventional neurofeedback system. Note the overall approach towards an optimal modulation value in the proposed system in contrast to the noisy, unpredictable progression in the neurofeedback system.

After trimming the first 15 seconds of acclimatizing to the situation, the slope for meditation score over time was computed for each participant in each feedback loop. This slope was interpreted as the goodness of the given system for the participant — a faster and more drastic improvement in meditation score is naturally interpreted as a better outcome. A processed dataset composed of the difference in the scores for each system for each participant was then created (Fig 7).

In order to test the hypothesis, a Z-score test was conducted on the processed dataset of differences between the scores

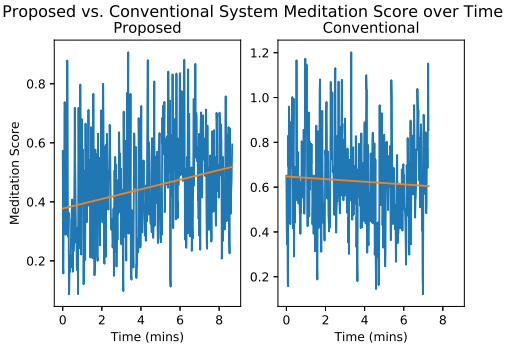

Fig. 7: Meditation scores over time with linear fit. Linear fit for score vs. time was employed to determine the effectiveness of the system.

for the conventional neurofeedback system and the proposed system. The null hypothesis H_0 was rejected with $p = 0.06$, indicating a strong chance that the novel neurofeedback system is better for achieving a meditative brain state in our experimental setup.

B. Discussion

Due to COVID-19 constraints, only 10 total data points were collected to form these conclusions. A greater sample size would certainly increase the confidence in these findings. As well, control in the wakefulness of participants is of interest as several participants noted their fatigue interfering with their ability to meditate. A greater trial length would similarly help in increasing confidence in the findings, and exploration of further audio modulation techniques is warranted to determine the optimal configuration.

IV. Conclusions

We proposed a novel approach to neurofeedback where the user's EEG data is passed first to a deep reinforcement learning agent that process the information using a user-specified reward function, that in turn controls an audio modulation system. It was shown that, in the case of optimizing for mindfulness and meditation EEG correlates, the proposed system outperformed the conventional neurofeedback system with $p = 0.06$.

Overall, this new paradigm for interpreting BCI information in a Humanistically Intelligent manner clearly holds promise for improvement of our practical understanding of the mind's idiosyncratic reaction to changes in a stimulus. Areas for further inquiry include longer-term experimentation where the same model is trained for multiple sessions in a row for a given user. Non-specific modeling for human response to modulation (i.e. the same model trained on multiple participants) would also be of interest. Applications in areas other than music therapy and meditation are also worthy of exploration.

V. Acknowledgements

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC).

REFERENCES

[1] J. Cohen, "Electroencephalography," 2014. [Online]. Available: www.accessscience.com

[2] N. Burrous, *Standard EEG: A Research Roadmap for Neuropsychiatry*.

[3] S. Mann, P. V. Do, D. E. Garcia, J. Hernandez, and H. Khokhar, "Electrical engineering design with the subconscious mind," in *IEEE International Conference on Human-Machine Systems (ICHMS)*, to be published, 2020.

[4] S. Enriquez-Geppert, R. J. Huster, and C. S. Herrmann, "Eeg-neurofeedback as a tool to modulate cognition and behavior: A review tutorial," *Frontiers in Human Neuroscience*, vol. 11, no. 51, Feb 2017.

[5] R. Ramirez, M. Palencia-Lefler, S. Giraldo, and Z. Vamvakousis, "Musical neurofeedback for treating depression in elderly people," *Frontiers in Neuroscience*, vol. 9, no. 354, Oct 2015.

[6] S. Mann, "Humanistic intelligence/humanistic computing: 'wearcomp' as a new framework for intelligent signal processing," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2123–2151+cover, Nov 1998.

[7] ——, "Wearable computing: Toward humanistic intelligence," *IEEE Intelligent Systems*, vol. 16, no. 3, pp. 10–15, May/June 2001.

[8] ——, "Phenomenological augmented reality with the sequential wave imprinting machine (swim)," in *2018 IEEE Games, Entertainment, Media Conference (GEM)*. IEEE, 2018, pp. 1–9.

[9] P. Scourboutakos, M. H. Lu, S. Nerker, and S. Mann, "Phenomenologically augmented reality with new wearable led sequential wave imprinting machines," in *Proceedings of the Tenth International Conference on Tangible, Embedded, and Embodied Interaction*. ACM, 2017, pp. 751–755.

[10] M. Irrmischer, S. J. Houtman, H. D. Mansvelder, M. Tremmel, U. Ott, , and K. Linkenkaer-Hansen, "Controlling the temporal structure of brain oscillations by focused attention meditation," *Human Brain Mapping*, vol. 39, 2018.

[11] R. Liu and J. Zou, "The effects of memory replay in reinforcement learning," in *2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2018, pp. 478–485.

[12] B. Mavrin, S. Zhang, H. Yao, L. Kong, K. Wu, and Y. Yu, "Distributional reinforcement learning for efficient exploration," *CoRR*, vol. abs/1905.06125, 2019. [Online]. Available: http://arxiv.org/abs/1905.06125

[13] C. J. C. H. Watkins, "Learning from delayed rewards," Ph.D. dissertation, Royal Holloway University of London, London, 1989.

[14] A. K. Dixit and J. J. F. Sherrerd, *Optimization in Economic Theory*. Oxford: Oxford University Press, 1990.

[15] V. Minh, K. Kavukcuoglu, and D. Silver, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, pp. 529–533, Feb 2015.